



FACULTÉ DE MÉDECINE
DE MARSEILLE

UNIVERSITÉ DE LA MÉDITERRANÉE
AIX-MARSEILLE III

TP STA-MQGR

UTILISATION PRATIQUE DU LOGICIEL STATISTIQUE « R »

Seydou Barro

SESSTIM, UMR 912, Aix-Marseille Université

golo-seydou.barro@univ-amu.fr

INTRODUCTION

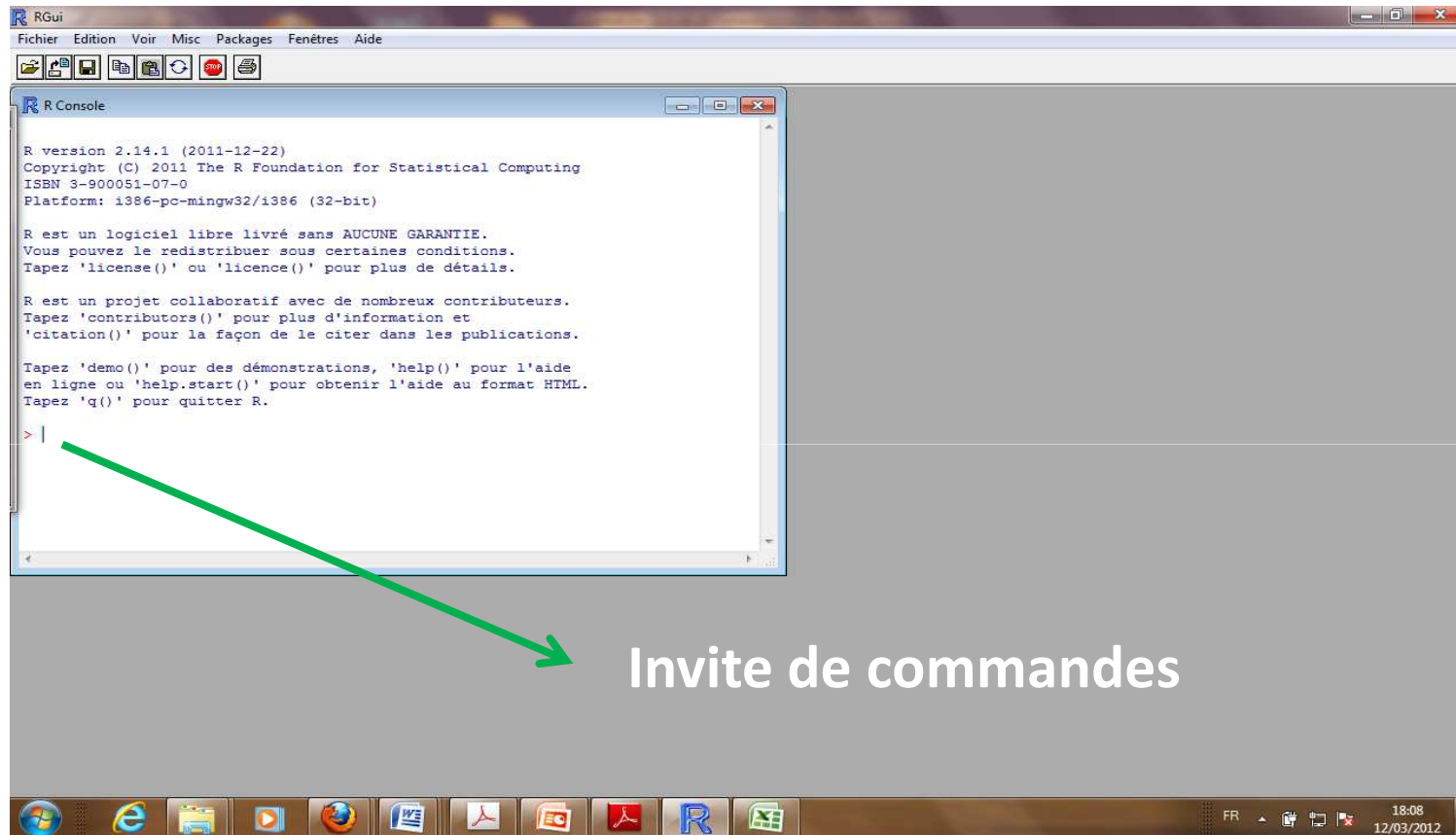
- R est un logiciel libre distribué gratuitement
- Il est très puissant et fournit les procédures usuelles pour l'analyse statistique des données
- Il possède des possibilités graphiques performantes pour explorer les données
- Il peut être utilisé aussi bien en mode interactif qu'en mode batch
- Il est multiplateforme (utilisable sur Linux, Windows et Mac OS)

DECOUVERTE DE R

Installation de R

- Site principal: www.r-project.org
- Le téléchargement: à partir d'un des sites du ``*Comprehensive R archive Network*'' (CRAN)
- Téléchargement et installation de la version récente compatible avec l'environnement de votre ordinateur
 - [Download R for Linux](#)
 - [Download R for \(Mac\) OS X](#)
 - [Download R for Windows](#)

LANCEMENT DU LOGICIEL



Environnement de travail: Console

Premiers essais sur R

R Comme calculatrice:

```
> 4+4
```

```
[1] 8
```

Notion d'objet:

```
> x<-5 : (X = l'objet, 5= la valeur de l'objet)
```

```
> x
```

```
[1] 5
```

Notion de vecteur:

- Ensemble de valeur contenues dans un opérateur
- Exemple de relevés de chiffres de la TA:

```
> TA<-c(10,11,8,13,15,11,16,9,19,26)
```

```
> TA
```

```
[1] 10 11 8 13 15 11 16 9 19 26
```

C = Vecteur

10, 11,8,..... = Valeurs du vecteur

Saisir des données dans R

- On peut saisir des données directement dans R à partir du clavier.
- On peut le faire avec les commandes **scan()** ou **data.frame()**
- **Exemple**: on dispose de deux variables, taille en cm et poids en kg, mesurées chez 6 individus
 - taille : 176 ; 185 ; 162 ; 188; 155
 - poids : 76 ; 65 ; 79 ; 84; 58

Saisir des données dans R

- Avec la commandes **scan()**

```
> taille <-scan ()
```

```
176 185 162 188 155
```

```
> poids <-scan ()
```

```
76 65 79 84 58
```

```
>donnees <-data.frame (taille, poids)
```

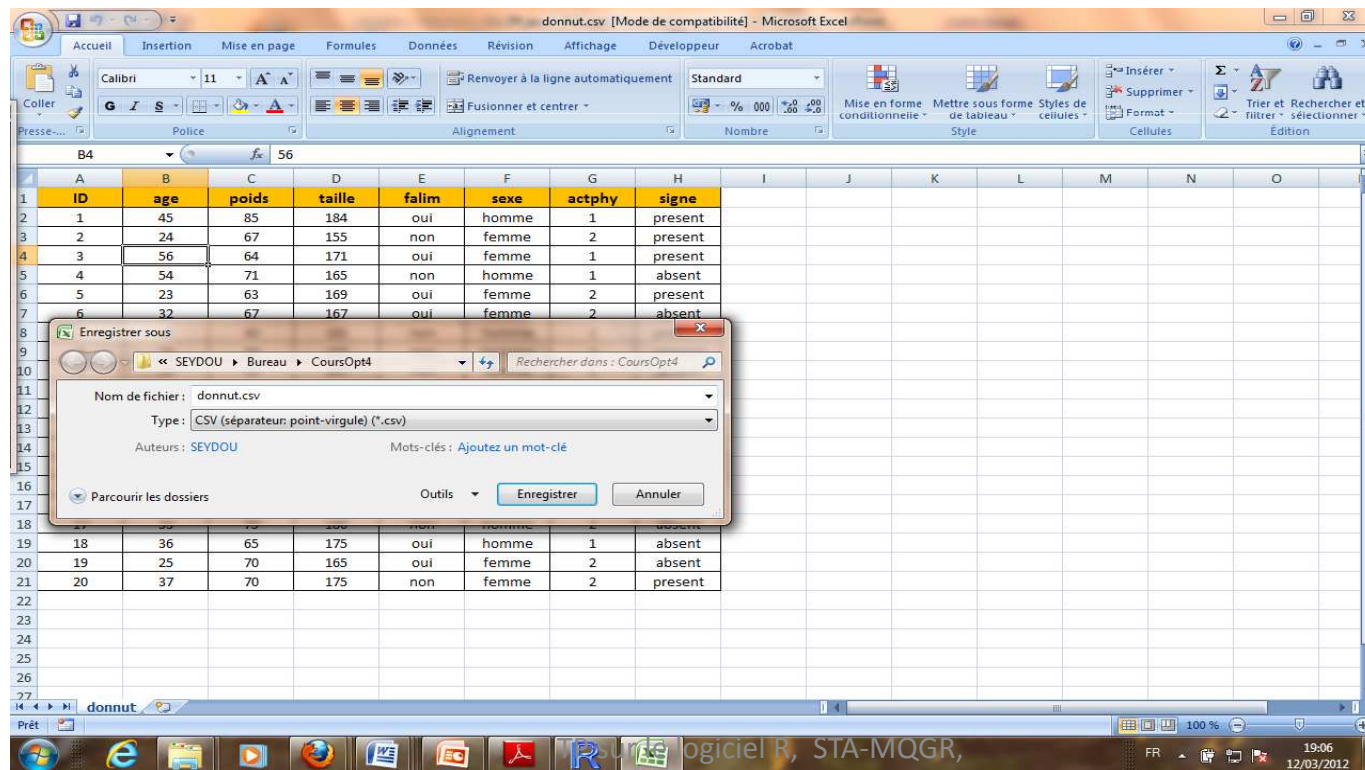
```
> edit (donnees)
```

- Avec **data.frame ()** : **data.frame()** fait appel à la commande **c()** :

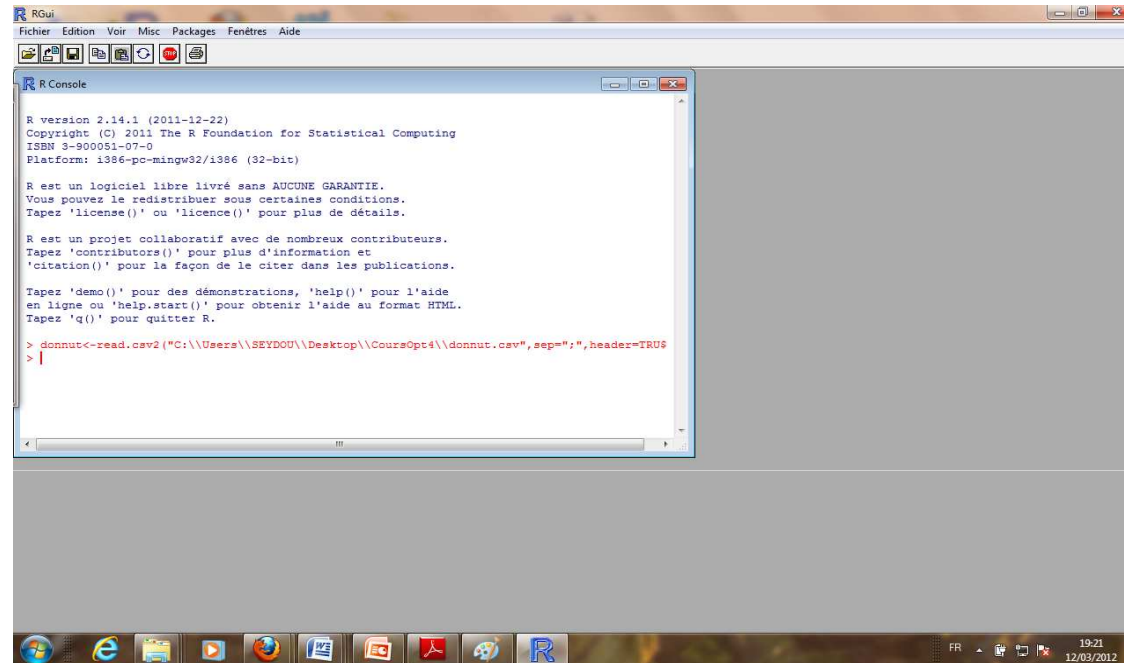
```
> data.frame(taille=c(176 185 162 188 155), poids=c(76 65 79 84 58))
```

UTILISATION DES DONNEES

- Utilisation de données externes:
- Importation de « Fichier Excel »
- Sauvegarde du fichier sous format .csv



Lire le fichier excel.csv avec R



```
>donnut<-read.csv2("C:\\Users\\SEYDOU\\Desktop\\STA-MQGR\\donnut.csv",sep=";",header=TRUE)
>
```

Vérification de l'intégrité des données

The screenshot shows the RGui interface with the R Console displaying the following code and output:

```
> donnut<-read.csv2("C:\\Users\\SEYDOU\\Desktop\\CoursOpt4\\donnut.csv", sep=";", header=TRUE)
> donnut
```

	ID	age	poids	taille	falim	sexe	actphy	signe
1	1	45	85	184	oui	homme	1	present
2	2	24	67	155	non	femme	2	present
3	3	56	64	171	oui	femme	1	present
4	4	54	71	165	non	homme	1	absent
5	5	23	63	169	oui	femme	2	present
6	6	32	67	167	oui	femme	2	absent
7	7	32	80	185	non	homme	1	present
8	8	26	69	155	non	homme	2	present
9	9	26	55	167	non	femme	1	absent
10	10	60	81	181	oui	homme	1	absent
11	11	39	74	168	oui	femme	1	absent
12	12	41	54	154	non	femme	2	absent
13	13	46	73	172	oui	homme	1	absent
14	14	43	65	164	non	femme	1	present
15	15	57	74	180	oui	homme	2	present
16	16	36	79	175	oui	homme	2	present
17	17	33	75	180	non	homme	2	absent
18	18	36	65	175	oui	homme	1	absent
19	19	25	70	165	oui	femme	2	absent
20	20	37	70	175	non	femme	2	present

Annotations in the image:

- A red arrow points from the word "Colonnes" to the header row of the data table.
- A green bracket on the left side of the data rows is connected by a red arrow to the word "Lignes".

STATISTIQUES DE BASE

- Commande **attach()**
- Permet de manipuler les variables facilement

```
>donnut<-read.csv2("C:\\Users\\SEYDOU\\Desktop\\ STA-MQGR \\donnut.csv",sep=";",header=TRUE)
```

```
>attach(donnut)
```

```
>poids
```

```
[1] 85 67 64 71 63 67 80 69 55 81 74 54 73 65 74 79
```

```
[17] 75 65 70 70
```

STATISTIQUES DE BASE

> mean(age) = Age moyen
[1] 38.55

> median(poids) = Poids médian
[1] 70

> var(taille) = Variance des tailles
[1] 86.02895

> sd(taille) = Ecart-type des tailles
[1] 9.275179

> min(taille) = Taille minimale
[1] 154

> max(taille) = Taille maximale
[1] 185

STATISTIQUE DE BASE

Fonction **summary()**

- `summary()` donne les statistiques élémentaires
- `>summary(donnut)` = commentaires

Commande **tapply()**

- Stratifie les statistiques par groupe
- Exple: statistiques élémentaires par sexe:
- `>tapply(poids,sexe,summary)` = commentaires

Commande **subset()** selectionne un groupe

Exple: données concernant les hommes

- `> subset(donnut,sexe=="homme")` = commentaire

TEST STATISTIQUES COURANTS

(tests paramétriques)

Comparaison de deux moyennes

- **Cas 1:** Variance homogènes (test de Student)
- `>t.test(taille[sexe=="homme"],taille[sexe=="femme"],var.equal=TRUE)`
- **Cas 2:** Variance non homogène (test d'Aspin-welch)
- `>t.test(taille[sexe=="homme"],taille[sexe=="femme"])`
- Commentaires et différence entre ces 2 tests

TESTS STATISTIQUES

- **Comparaison de deux variances**
- Test de Bartlett
- Dispersion de la taille par sexe
- `> bartlett.test(taille,sexe)`
- **Comparaison de deux pourcentages**
- Test du Khi-deux
- Comparer le % de malades en fonction du sexe:
- `>chisq.test(table(sexe,signe))`

TESTS NON PARAMETRIQUES

- Pour plus de 2 échantillons indépendants
- `>kruskal.test(taille,sexe)`
- **Test de conformité** (test de shapiro-wilks)
- Tester si un échantillon provient bien d'une loi de probabilité donnée
- Test de kolmogorov-Smirnov (KS)
- Peut être utilisé aussi pour d'autres lois
- `> ks.test(taille,"pnorm",mean(taille),sd(taille))`
- `>shapiro.test(taille)`

Quelques graphiques avec R

Nuage de points

- Taille en fonction du poids
- `> plot(taille, poids,xlab="taille(en cm)",ylab="poids(en kg)",col="blue",main="taille en fonction du poids")`

Nuage de points avec droite de régression

- `> plot(poids~taille)`
- `> abline(lm(poids~taille))`

Boxplot (boîte à moustaches)

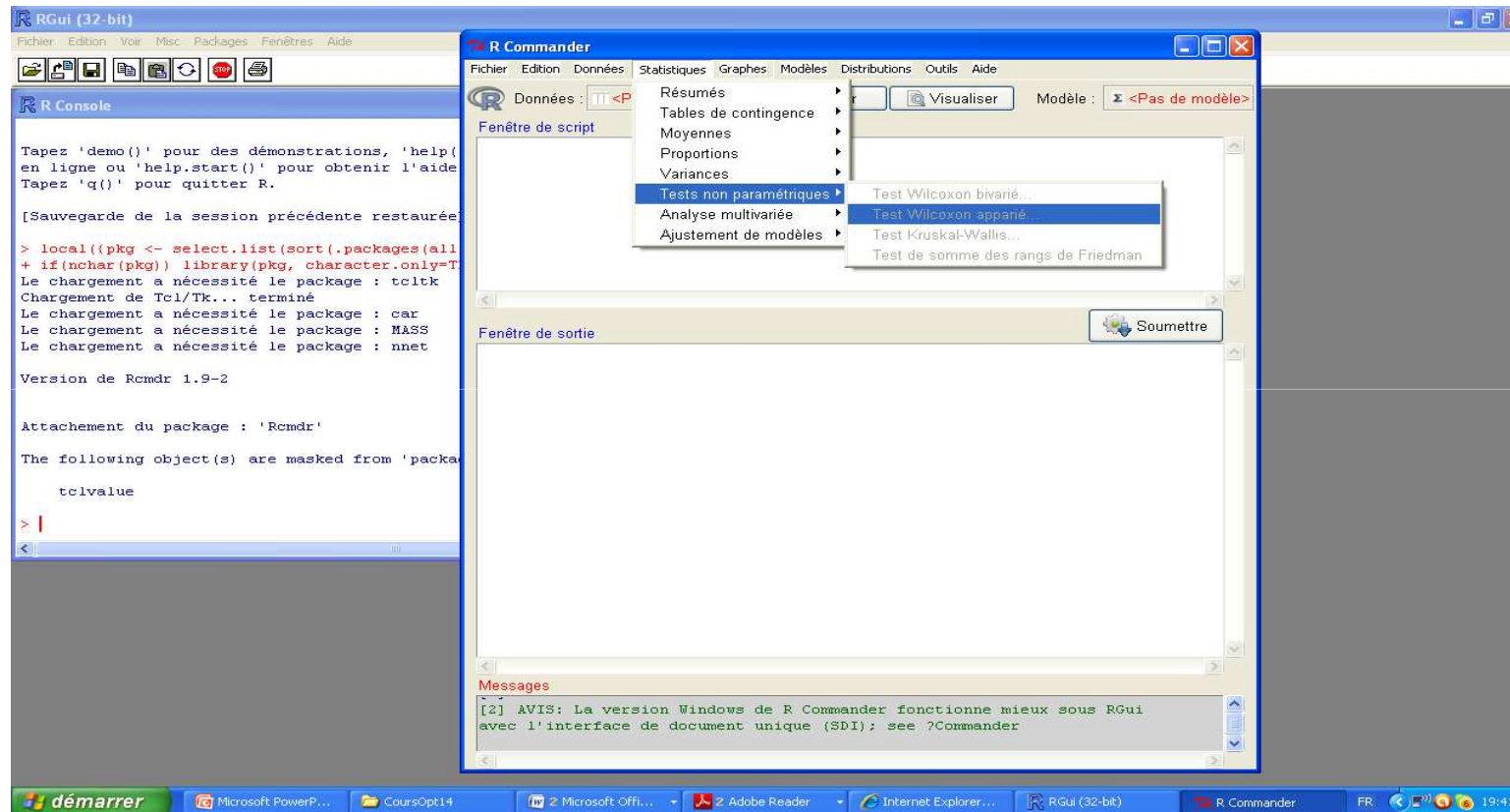
- `> boxplot(taille,horizontal=TRUE,col="lightblue",main="boxplot des tailles")`

Boxplot par groupe

- `> boxplot(taille~sexe,col="lightblue",main="box plot des tailles en fonction du sexe")`

R COMMANDER: package Rcmdr

Interface graphique de R



Sans programmation par ligne de commande

Quelques liens utiles

Ressources sur le Web

- <http://www.r-project.org/>
- <http://cran.r-project.org/manuals.html>
- <http://cran.r-project.org/doc/manuals/R-intro.html>
- <http://cran.r-project.org/doc/manuals/R-data.html>
- <http://cran.r-project.org/faqs.html>
- <http://wiki.r-project.org/>
- <https://stat.ethz.ch/mailman/listinfo/r-announce>
- <http://journal.r-project.org/>
- <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>
- <http://rforsasandspssusers.com/>
- <http://www.statmethods.net/>
- <http://www.jstatsoft.org/>
-
- **Ressources francophones**
- http://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf
- <http://pbil.univ-lyon1.fr/R/enseignement.html>
- <http://pbil.univ-lyon1.fr/R/html/cours1>
- http://zoonek2.free.fr/UNIX/48_R_2004/all.html
- <http://www.r-project.org/mail.html>